INSIGHTS · INSIGHT QUALITY AND PITFALLS

Ethical Use of Consumer Insight

The ethical questions raised by collecting and using knowledge about consumers, including consent, privacy and context, re-identification, sensitive inferences, manipulation and unfair treatment, with practical safeguards for insight teams.

ModQu research article · 1,580 words · 18 references (APA 7th edition)

Consumer insight exists to help organisations understand people well enough to serve them better. The same understanding can be used to exploit them. Knowing that a customer is anxious, short of money or about to make a major life change can help an organisation offer the right support, or help it press for a sale at a vulnerable moment. As organisations gather more data and use more powerful methods to analyse it, the ethical questions become harder to avoid. This article sets out the main questions and the safeguards that help insight teams answer them.

Why ethics is part of insight quality

Ethics is sometimes treated as a legal constraint added at the end of a project. It is better understood as part of the quality of insight work. Research that participants would object to if they understood it damages trust, and trust is what makes people willing to answer surveys, share data and buy from an organisation. Martin and Murphy (2017) reviewed research on data privacy in marketing and concluded that consumers' concerns about how their data are used affect their trust in firms and their willingness to engage with them. An insight obtained in a way that customers would regard as a betrayal can cost more than it is worth.

The research professions have long recognised this. The Belmont Report set out three principles for research with human participants: respect for persons, which requires informed consent; beneficence, which requires minimising harm; and justice, which requires a fair distribution of the burdens and benefits of research (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). Market research codes apply similar ideas. The ICC/ESOMAR (2016) code requires researchers to be honest with participants, to respect their right to withdraw, to protect their personal data and to keep research clearly separate from selling.

Consent and its limits

Most data used for consumer insight are collected with some form of consent, often by accepting terms and conditions. This consent is frequently thin. McDonald and Cranor (2008) estimated that reading the privacy policies of the websites a typical person visits would take each American about 200 hours a year. Obar and Oeldorf-Hirsch (2020) asked participants to join a fictitious social network and found that most skipped the privacy policy and terms of service, and almost all agreed to terms that included sharing their data with employers and giving up their first-born child. Consent obtained this way is a legal fact but weak evidence that people understand or accept how their data will be used.

Acquisti et al. (2015) reviewed research on privacy behaviour and showed that people's privacy choices are uncertain, depend strongly on context and can be easily influenced by how options are presented. They concluded that leaving privacy entirely to individual choice does not adequately protect people.

Context: would people expect this use?

A more useful question than "did they consent?" is "would they expect this?". Nissenbaum (2004) proposed that privacy is best understood as contextual integrity: information flows carry expectations attached to the context in which they occur, and a privacy violation occurs when information moves in ways that break those expectations. Purchase data shared with a retailer to process an order carry different expectations from the same data sold to an insurer. Solove (2006) set out a taxonomy of privacy problems, distinguishing harms arising from information collection, processing, dissemination and intrusion, and showed that many privacy harms arise not from collection itself but from aggregation and secondary use.

Five questions before using consumer data

A "no" or "not sure" to any of them is a reason to pause

Question	Ask
1. Consent	Did people knowingly agree to this use?
2. Context	Would they expect their information to be used this way?
3. Sensitivity	Does it reveal or infer health, money, beliefs or other sensitive traits?
4. Harm	Could the use disadvantage, exclude or exploit anyone?
5. Transparency	Could we explain this use publicly without embarrassment?

Based on Nissenbaum (2004), Solove (2006) and ICC/ESOMAR (2016). Diagram by ModQu ModQu.com

Figure 1. Five questions to ask before using consumer data or insight. A "no" or "not sure" to any of them is a reason to pause. Diagram by ModQu, based on Nissenbaum (2004), Solove (2006) and the ICC/ESOMAR (2016) code.

Re-identification and sensitive inferences

Removing names does not make data anonymous. Sweeney (2002) showed that a large share of the United States population could be uniquely identified by the combination of postcode, date of birth and sex, and used this to re-identify individuals in supposedly anonymous health records. Narayanan and Shmatikov (2008) showed that individuals in a large anonymised dataset of film ratings could be identified by matching a few of their ratings with public information. Insight teams that share or publish customer-level data should assume that re-identification is possible unless strong protections are in place.

Analysis can also reveal what people never disclosed. Kosinski et al. (2013) showed that records of what people "liked" on a social network could be used to predict, with considerable accuracy, sensitive attributes including sexual orientation, political views, religion and personality. A widely reported case described how a retailer's analysis of purchases could indicate that a customer was likely to be pregnant, raising questions about whether customers would welcome being marketed to on that basis (Duhigg, 2012). Inferences of this kind can be more intrusive than direct questions, because the person does not know the inference has been made.

Manipulation and exploitation

Insight into how people make decisions can be used to help them decide well or to steer them against their interests. Susser et al. (2019) defined online manipulation as hidden influence that exploits people's vulnerabilities to steer their decisions, and argued that it is wrong because it undermines autonomy, even when it does not cause obvious harm. Mathur et al. (2019) examined around 11,000 shopping websites and found more than 1,800 instances of "dark patterns", interface designs that push users towards choices they might not otherwise make, such as false scarcity messages, hidden costs and obstacles to cancelling.

Experiments on customers raise similar issues. Kramer et al. (2014) altered the emotional content of the news feeds of about 690,000 social network users to study emotional contagion. The study prompted widespread criticism, because users had not knowingly agreed to have their emotions experimentally manipulated. Routine testing of designs and offers is a normal part of insight work, but tests that aim to change people's emotional states or exploit their weaknesses need a higher bar of justification.

Fairness

Insights and models built from historical data can reproduce historical unfairness. Barocas and Selbst (2016) explained how data mining can produce discriminatory results without any intent to discriminate, because the data reflect past discrimination or because variables that seem neutral act as proxies for protected characteristics. Obermeyer et al. (2019) found that a widely used algorithm for allocating extra health care to patients underestimated the needs of Black patients, because it used past health spending as a measure of need and less had historically been spent on Black patients with the same level of illness. A pricing, targeting or eligibility rule derived from customer insight should be checked for its effects on different groups before it is used.

Regulation sets the floor

Data protection law sets minimum standards. The European Union's General Data Protection Regulation requires, among other things, a lawful basis for processing personal data, that data be collected for specified purposes and not used in ways incompatible with them, that collection be limited to what is necessary, and that special protection be given to sensitive categories such as health and religion (Regulation (EU) 2016/679, 2016). Legal compliance is necessary but not sufficient. Many uses that are legal would still surprise or upset customers if they knew of them.

Practical safeguards

Several practices help insight teams use consumer knowledge ethically:

- **Ask the five questions** in the figure above before starting a project that uses personal data or that could affect how customers are treated.
- **Collect only what is needed.** Data that are not collected cannot be misused or leaked.
- **Separate research from selling.** Keep data collected for research out of individual targeting unless participants agreed to that use.
- **Protect identities.** Aggregate results, remove or alter identifying details, and restrict access to customer-level data.
- **Be careful with inferences.** Treat inferred sensitive traits with the same care as disclosed ones, or avoid making them.
- **Check for unequal effects.** Test models and rules for differences in outcomes across groups.
- **Apply the publicity test.** If the organisation would be uncomfortable explaining a use of data to the people concerned, it should reconsider.

Some organisations also set up a simple review step for projects that pass a threshold, such as those that link data sources, infer sensitive traits, run experiments on customers or feed automated decisions. A small group that includes people from outside the insight team reads a short description of the project and asks the five questions. The aim is not to block work but to catch problems while they are still cheap to fix, and to create a record of why a use of data was judged acceptable. Particular care is needed with vulnerable groups, such as children, people in financial difficulty and people who are ill or bereaved, because the same insight that helps an organisation support them can also be used to take advantage of them.

Key takeaways

Consumer insight can be used to serve people or to exploit them. Ethical use is part of insight quality, because trust underpins customers' willingness to share information and to buy. Formal consent is often thin, so it is better to ask whether people would expect a particular use of their information. Removing names does not guarantee anonymity, and analysis can infer sensitive traits people never disclosed. Insight into decision making can slide into manipulation, and models built from historical data can reproduce unfairness. Law sets a minimum standard; ethical practice asks more, through clear questions and simple safeguards applied before data are used.

References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509–514. https://doi.org/10.1126/science.aaa1465

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671–732. https://doi.org/10.15779/Z38BG31

Duhigg, C. (2012, February 16). How companies learn your secrets. *The New York Times Magazine*.

ICC/ESOMAR. (2016). *ICC/ESOMAR international code on market, opinion and social research and data analytics*. ESOMAR.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences, 110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, 111*(24), 8788–8790. https://doi.org/10.1073/pnas.1320040111

Martin, K. D., & Murphy, P. E. (2017). The role of data privacy in marketing. *Journal of the Academy of Marketing Science, 45*(2), 135–155. https://doi.org/10.1007/s11747-016-0495-4

Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction, 3*(CSCW), Article 81. https://doi.org/10.1145/3359183

McDonald, A. M., & Cranor, L. F. (2008). The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society, 4*(3), 543–568.

Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In *Proceedings of the 2008 IEEE Symposium on Security and Privacy* (pp. 111–125). IEEE. https://doi.org/10.1109/SP.2008.33

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. U.S. Department of Health, Education, and Welfare.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review, 79*(1), 119–157.

Obar, J. A., & Oeldorf-Hirsch, A. (2020). The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society, 23*(1), 128–147. https://doi.org/10.1080/1369118X.2018.1486870

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

Regulation (EU) 2016/679. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union, L 119*, 1–88.

Solove, D. J. (2006). A taxonomy of privacy. *University of Pennsylvania Law Review, 154*(3), 477–564. https://doi.org/10.2307/40041279

Susser, D., Roessler, B., & Nissenbaum, H. (2019). Online manipulation: Hidden influences in a digital world. *Georgetown Law Technology Review, 4*(1), 1–45.

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(5), 557–570. https://doi.org/10.1142/S0218488502001648